The main statistical tool used in this project was logistical regression on the Viral Loads and CD4 counts in the HIV data set. The data was subjected to a 80/20 split for training and testing purposes.

**Result 1 – Confusion Matrix**

A Confusion Matrix was applied to the data 15 times and the results were logged to produce means for the TP,FP,FN,TN. The resulting process produced:

|  |  |  |
| --- | --- | --- |
|  | **0** | **1** |
| **0** | 151.200 | 34.733 |
| **1** | 7.067 | 7.000 |

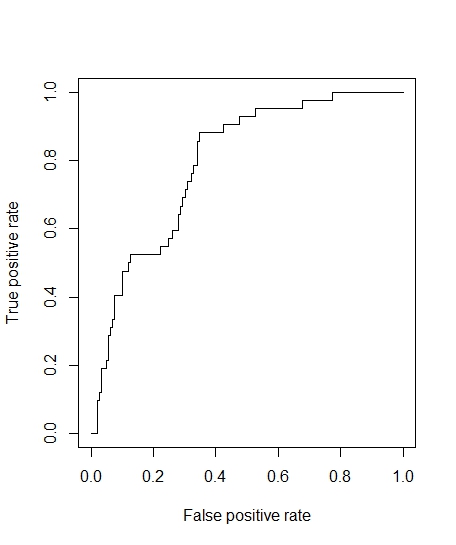
Accuracy rate = 79.1%

Precision = 95.5%

Recall = 81.3%

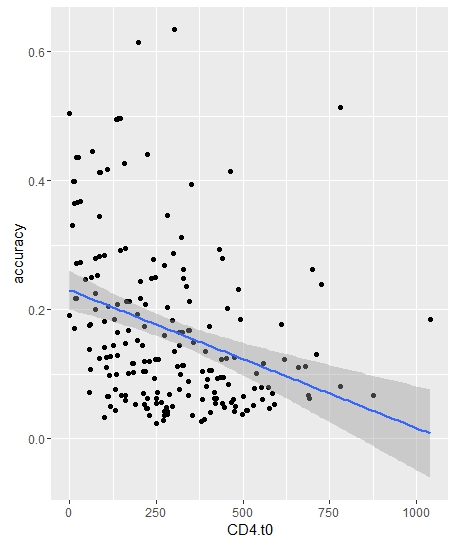
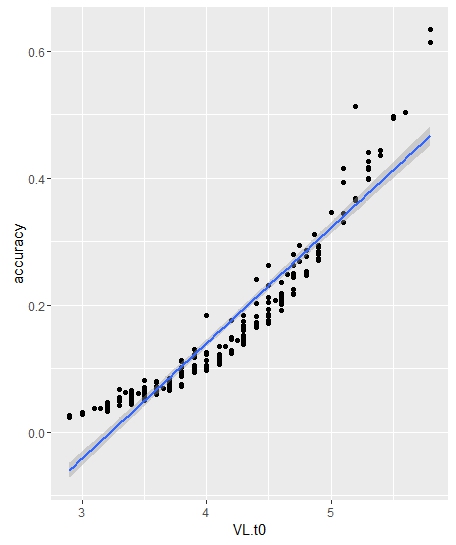
**Result 2 – ROC Plot**

As we can see, the curve follows fairly close to the left-hand border and top border which indicates it is an accurate test.



**Result 3 – Slope**

As we can tell from the following two visuals, Viral Load is a much better predictor of the response than White Blood Cells.



**Result 4 – P-Value**

The p-value for CD4 (White Blood Cells) is 0.3341 which naturally means it is not a good predictor. However, the p-value for VL (Viral Load) is a great predictor with the tiny p-value of 2e-16.

